

DERNIÈRES AVANCÉES DANS LE SÉQUENÇAGE DU GÉNOME DE *THEOBROMA CACAO* CRIOLLO : NOUVEL ASSEMBLAGE, ANNOTATION ET OUTILS.

Xavier Argout¹, Jean-Marc Aury^{2,3,4}, Mark Guitinan⁵, Gaetan Droc¹, Jerome Gouzy⁶, Siela Maximova⁵,
Olivier Fouet¹, Stephan Schuster⁷, Michel Boccara^{1,8}, Didier Clement^{1,9}, Mathias Tahi¹⁰,
Flavine Gramachet⁵, Claire Lanaud¹

¹ CIRAD, UMR 1145 C4D3, TA 100, 17170 Saint-Pierre, Réunion
² CEA, ICG, Genoscope, 2 rue Gaston Crémieux, 91191 Evry, France
³ CNRS, UMR 8030, 2 rue Gaston Crémieux, CP5706, F-91057 Evry, France
⁴ Université d'Evry, F-91057 Evry, France

⁵ Penn State University, Department of Horticulture and the Huck Institutes of the Life Sciences, University Park, PA 16802, États-Unis

⁶ INRA-CNRS LIPM Laboratoire des Interactions Plantes Micro-organismes, BP 52627, 31326 Castanet Tolosan Cedex, France

⁷ Penn State University, Department of Biochemistry and Molecular Biology, University Park, PA 16802, États-Unis

⁸ University of the West Indies, Cocoa Research Unit (CRU), St Augustine, Trinité et Tobago

⁹ Comissão Executiva de Planejamento da Lavoura Cacaueira (CEPLAC), Km 22 Rod. Ilheus Itabuna, Cx. postal 07, Itabuna 45600-00, Bahia, Brésil

¹⁰ Centre National de la Recherche Agronomique (CNRA), B.P. 808, Divo, Côte d'Ivoire

RESUME

Le 26 décembre 2010, le Consortium international pour le séquençage du génome du cacaoyer (International Cocoa Genome Sequencing Consortium -ICGS) a publié dans *Nature Genetics* la première séquence du génome de *Theobroma cacao* L. Le génotype choisi, le B97-61/B2, est une variété de Criollo qui apporte un arôme fin au chocolat. Cette première séquence comportait 76 % de la couverture du génome et 67 % de l'assemblage a été ancré dans une carte génétique à haute densité. Au cours de l'année qui vient de s'écouler, de nouveaux progrès ont été faits sur l'assemblage, l'annotation du génome et l'identification de familles de gènes. De plus, de nouveaux outils ont été développés pour mieux exploiter la séquence. Un navigateur de génome a été créé pour permettre aux utilisateurs d'accéder à la séquence, de visualiser leurs annotations (structure des gènes, fonction putative du gène, analyse génomiques comparatives, expression des gènes dans différents tissus/conditions, etc.) et interroger les données avec plusieurs sources biologiques, effectuer une comparaison de séquence et lancer une recherche par mots-clés. L'assemblage actuel représente 87,1 % de la taille estimée du génome (430 Mb). Cet assemblage semble couvrir une très large proportion de l'euchromatine du génome de *T. cacao*, permettant de récupérer 97,8 % des ressources sur les unigènes (38 737 unigènes assemblés à partir de 715 457 séquences EST) dans l'assemblage du génome. Les annotations ont révélé 28 798 gènes codants de protéines parmi lesquels 86,1 % pourraient être ancrés dans une carte génétique à haute densité. Seulement 20 % du génome se composait d'éléments transposables, soit un pourcentage significativement inférieur à comparer à d'autres génomes de taille similaire. Cette séquence de génome va faciliter une meilleure compréhension de l'élaboration des caractéristiques et accélérera la sélection de *T. cacao* par une sélection assistée par marqueurs efficace et l'exploitation des ressources génétiques. Le navigateur de génome est disponible gratuitement sur le site web suivant : <http://cocoagendb.cirad.fr>.